Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Diplomová práce

Nástroj pro automatickou identifikaci KIR alel

Plzeň 2020 Kateřina Kratochvílová

Místo této strany bude zadání práce.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne 22. února 2020

Kateřina Kratochvílová

Poděkování

Ráda bych poděkovala Ing. Lucii Houdové, Ph.D. za cenné rady, věcné připomínky, trpělivost a ochotu, kterou mi v průběhu zpracování této práce věnovala.

Abstract

The text of the abstract (in English). It contains the English translation of the thesis title and a short description of the thesis.

Abstrakt

Text abstraktu (česky). Obsahuje krátkou anotaci (cca 10 řádek) v češtině. Budete ji potřebovat i při vyplňování údajů o bakalářské práci ve STAGu. Český i anglický abstrakt by měly být na stejné stránce a měly by si obsahem co možná nejvíce odpovídat (samozřejmě není možný doslovný překlad!).

Obsah

1 Úvod

2 Geny

V každé buňce lidského organismu, konkrétně v buněčném jádře, je možné nálest 46 chromozomů. Jeden chromozom představuje stočenou dlouhou molekulu DNA (Deoxyribonuklenovou kyselinu). Všech 46 chromozomů obsahuje okolo 100 000 genů. Drobný segment DNA, který řídí buněčnou funkci je právě gen. Konkrétní forma genu je alela. [9]

2.1 Nomenaklura

akorat jeste pred to by teda chtělo hodit jak vubec vypada genom

2.2 Imunitní systém

Imunitní systém chrání organismus před škodlivinami. Skládá se ze dvou hlavních částí vrozené imunity a získané imunity. Jiné označení pro vrozenou imunitu může být přirozená, neadaptivní nebo antigenně nespecifická. Jiné označení pro získanou imunitu je specifická nebo adaptivní. Pro tuto práci je důležitý fakt že NK buňky patří do přirozené imunity. NK buňky budou rozebírány dále v textu.

Vrozená imunita veškeré informace jsou neměnně zapsány v DNA - odpovídá po každíém setkání s antigenem stejným mechanismy nemá paměť - buňky se nechází neustále v kry a takže je aktivace v případě potřeby takřka okamžitá (minuty až hodiny)

Specifická imunita - v genomu jedince obsaženy pouze její základy - v průběhu vývoje dochází ke změnám genomu jednotlivých buněk, které se pak odraží na jejich fenotypu - specifická imunita se fyziologicky rozvíjí až po narození - nefunguje samostatně vždy spolupracuje s přirozenou imunitou

aktivace až po setkání se svým antigenem pomalejší nástup než nespecifické mechanismy jiný průběh u opakovaného setkání schopnost pamatovat si zdroj wikiskripta

Antigen jsou látky které imunitní systém rozpozná a reguje na ně. Antigen znamená cizorodá částice. Nejčastější antigeny jsou cizorodé látky z vnějšího prostředí. Antigeny z organismu samého nazýváme endoantigeny (endogenní antigeny). Alergen je exoantigen, který je u vnímavého jedince schopen vyvolat patologickou (alergickou) imunitní reakci.

Antigen prezentující buňky (APC) a MHC systém APC jsou buňky vlastního těla schopné fagocytovat (makrofágy, dendritické buňky, B-lymfocyty) – co pozřou, to naštípou na krátké peptidické sekvence a vystaví na svém povrchu k „posouzení“ kromě těchto „vzorků“ mají na povrchu i MHC molekuly (z angl. major histocompatibility complex) MHC jsou vysoce polymorfní a zcela specifické a unikátní pro každého jedince MHC určují individuální identitu všech tkání – proto můžou působit komplikace spojené např. s odvržením štěpu po transplantaci největší koncentrace MHC je v leukocytech, proto se u člověka používá spíše zkratka HLA (z angl. human leukocyte antigens) více o MHC najdete například na Wikipedii teprve komplex MHC molekuly s antigenem vystavený na povrchu buňky aktivuje příslušný T-lymfocyt

2.3 Imunitní systém, HLA a non-HLA geny

Human leucocyte antigen(HLA) je genetický systém, který je primárně zodpovědný za rozeznávání vlastního od cizorodého. Tento systém je složen právě z jednotlivých HLA genů rozpoznávající antigeny (cizorodé částice). Pokud HLA gen přijde do styku s antigenem je antigen zničen. HLA obsahuje pravděpodovně i geny odpovědné za intenzitu imunitní odpovědi.

HLA je rozsáhlý komplex genů, které determinují (určují, rozpoznávjí????) povrchové molekuly (antigeny) umístěné v plazmatické membráně buněk

Hlavní fyziologickou funkcí molekul MHC je předkládat antigeny nebo jejich fragmenty buňkám imunitního systému, především T-lymfocytům (prezentace antigenu je prvním předpokladem pro rozvoj imunitní reakce a tím obrany proti napadení mikroorganismy). Pomocí těchto molekul buňky imunitního systému vzájemně kooperují.

Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému.
základní rozdíl mezi HLA a non-HLA a kir

Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému. Z III třídy jsou to všechny, z II žádný a z I je to směs. Zjednodušeně můžeme říci, že geny které nejsou HLA jsou non-HLA. Tyto geny souvisejí též s funkcí imunitního systému, ne však vylučně s funkcí HLA.

2.3.1 Jak vypadá genom

Genová oblast HLA komplexu, se nalézá na krátkem raménku 6. chromozomu (6p21.31), zaujímá úsek dlouhý 3600 kb (3,6cM), tedy přibliţně jednu

tisícinu genomu. Obsahuje 224 genů; 128 funkčních genů a 96 pseudogenů a patří k regionům s nejvyšší genovou hustotou.

Uprostřed HLA oblasti se nachází úsek o velikosti 1 Mb, ve kterém bylo identifikováno na 70 genů, které se funkčně ani strukturně nepodobají HLA molekulám. Navzdory této skutečnosti se vžilo označení geny III. třídy, přičemž některé geny původně zařazené do této třídy jsou nověji označovány jako geny IV. třídy (viz. výše).

HLA-6.Chromozom a KIR 19.chromozom udíž se segregují nezávisle a HLA shodní dárci s příjemce mají obvykle různé složení KIR genů (Fryčová)

2.4 10/10

Ta je postavena na případě typizace 5 lokusů (HLA-A/B/C/DRB1/ /DQB1). Vstupním parametrem samotného vyhledávání je míra shody (match) či definované neshody (mismatch). Navržená metoda je platná nejen pro úplnou míru shody 10/10 (shoda HLA-A/B/C/DRB1/DQB1), ale i menší, např. 8/8 (HLA-A/B/C/DRB1), či požadovanou neshodu na konkrétních lokusech, např. 9/10 HLA-A mismatch.

Během vyhledávání se hodnotí shoda obou alel v lokusech HLA-A, HLA-B, HLA-C, HLA-DRB1 a HLA-DQB1. Cílem je najít dárce, který bude s příjemcem shodný v 10 znacích z 10. V závislosti na pacientově stavu a nízké pravděpodobnosti najít včas shodného nepříbuzného dárce je možné tolerovat odchylky v jednom nebo dvou znacích (9/10, 8/10). Každá odchylka však zvyšuje riziko rozvoje potransplantačních komplikací.
dědičnost HLA znaků
Každý člověk má tzv. fenotyp neboli soubor HLA znaků, který je složen právě ze dvou haplotypů. Každý z haplotypů je tvořen sadou antigenů obsahujících konkrétní alely. Polovinu těchto znaků zdědíme od matky a polovinu od otce. Z hlediska transplantace se v současné době považují za nejdůležitější (a proto se také nejpřesněji vyšetřují) HLA antigeny I. třídy A, B, C a antigeny II. třídy DR a DQ. Existuje ale řada dalších – tzv. minoritních antigenů, které dosud nejsou dostatečně probádány, a jejich vliv na průběh transplantace se teprve zkoumá. V současnosti je požadavek na míru shody 10/10 neboli v pěti HLA antigenech, konkrétně v (HLA -A, -B, -C, -DRB1 a –DQB1). Nejmenší možná shoda představuje 6/10 v genech (HLA -A, -B, -DRB1), ale zde bohužel pro pacienta vzniká smrtelné riziko odvržení štěpu.

Počet teoreticky možných kombinací HLA znaků u člověka dosahuje několika miliard. Je známo, že některé tkáňové typy (kombinace znaků) se vyskytují

v určitém národě či oblasti častěji, jiné jsou extrémně vzácné. Protože se jednotlivé znaky dědí, shodu mezi dvěma jedinci najdeme nejsnáze v pokrevním příbuzenstvu. Od rodičů na potomky se příslušná polovina znaků předává obvykle ve zmíněné kompletní sadě (haplotypu). Pro zjednodušení je uveden příklad, podle kterého je dle genetických zákonů možné dědit jednu ze čtyř možných variant výše zmíněných druhů HLA antigenů mezi sourozenci (obr. 3.2).

2.5 Natural Killer

Při transplantaci se mohou objevit reakce štěpu proti hostuteli nebo relaps onemocnění.(Návrat nemoci) Podle nedávných studii kromě HLA genů ovlivňují výsledky přijetí i non HLA geny jedním z nich může být i KIR (Kiler immunoglobulit like receptor tohle je asi blbě napsaný.) Jelikož jsou geny kódovány na různých chrmozomech (HLA 6 a KIR 19) takže HLA schodní dárci s příjemcem mají různé složení KIR genů. V případěch kdy je více schodných HLA dárců tak by se pak porovnávalo KIR.

Velká buňka imunitního systému, nepotřebuje antigen aby začala zabíjet. -nespecifická imunita - vrozená, neadaptivní - veškeré potřebné informace zapsaná v DNA. Odpovídá při každém setkání s antigenem stejně - nemá paměť -> tady si to pročiřečí KIR jsou receptory na povrchu NK buněk, NK zabíjejí na nazákladě interakce mezi kir receptorem a HLA molekulou na povrchu buňky

NK buňky maji shopnost identifikovat buňky vlastního MHC systému (HLA I.třídy) ktere jsou normálně exprimovány prakticky na všech buňkách v těle. Nádorové a některé virem napadené buňky potlačují expresi HLA I. třídy a tím se brání napadení cytotoxickými T lymfocyty. Molekuly HLA I. třídy rozpoznávají NK buňky pomocí pozitivních a negativních receptorů, které mohou inhibovat nebo naopak aktivovat NK buňky k „zabíjení“

V užším slova smyslu se jako ligand označuje signální molekula, která se váže na vazebné místo cílového proteinu. Ligand, který je schopný po navázání na receptor vyvolat fyziologickou odpověď, se nazývá agonista, ten, který je schopen se vázat, ale odpověď nespouští, je antagonista

Zjednodušeně: NK buňky neustále systematicky zjišťují přítomnost či absenci příslušný HLA ligand pro své KIR receptory. Pokud je HLA molekula přítomna, pak dojde k vazbě KIR-ligand HLA a protože za normální okolností převládají inhibiční KIR nad aktivačními, tak nedojde ke spuštění cytoxické reakce NK buněk. Jestliže receptory KIR nenalzenou příslušný ligand HLA (vlastní molekulu HLA) aktivační KIR receptory převládnou nad

inhibičními a je spuštěna náležitá cytoxická kaskáda. lymfocyty bílá krvinka je leukocyt - typ bbílé krvinyky - T a B lymfocyty - specifická imunita - NK buňky nespecifická imunita - vznikají v z lymfatických kmenových buněk v kostní dřeni Aha takže lymfatické řečiště je více propustné proto to co nejde do cév jde sem pak se to odfiltruje a pak se to vrací do krevního řečiště.

KIR jsou na povrchu NK buněk a kde jsou teda NK buňky? NK je v podstatě lymfocyt a to je typ bílé krvinky. jo a nebudou teda spíš v lymfatické uzlině? leukocyty 1. granulocyty - neutrofilní, bazofilní a eozinogilní 2. agranulocyty - lymfocyty a monocyty

neutrofilní granulocyty jsou schopny vycestovat z kapilár do místa zánětu přeměněné monocyty přítomné v játrech v tělních dutinách (hrudní, břišní), ve slezině vy lymfatických uzlinách a kostní dřeni

2.6 Jak funguje HLA

2.7 Jak funguje non-HLA

2.8 Bordel pro prvni kapitolu

Takže to vypadá že nejdřív se najde shoda HLA a pak se ještě dodělávár KIR shoda. Proč KIR? pOrotože roste počet důkazů vluvu genů KIR že mají vliv na výsledky transplance při leukemii HLA je na 6. chormozomu KIR je 19 chomozomu. tudiž se segregují nezávisle a Hla shodní dárci s příjemcem mají obvykle různé složení KIR genů Nesmírná variabilita alel tohoto systému ztěžuje úspěšnost allogeních transplantací.

HLA jen zkopírováno a je ta i hezkej obrázek z Genová oblast HLA komplexu, se nalézá na krátkem raménku 6. chromozomu (6p21.31), zaujímá úsek dlouhý 3600 kb (3,6cM), tedy přibliţně jednu tisícinu genomu. Obsahuje 224 genů; 128 funkčních genů a 96 pseudogenů a patří k regionům s nejvyšší genovou hustotou.

Uprostřed HLA oblasti se nachází úsek o velikosti 1 Mb, ve kterém bylo identifikováno na 70 genů, které se funkčně ani strukturně nepodobají HLA molekulám. Navzdory této skutečnosti se vţilo označení geny III. třídy, přičemţ některé geny původně zařazené do této třídy jsou nověji označovány jako geny IV. třídy (viz. výše).

HLA nomenklatura HLA nomenklatura - zase jen skopírováno Vysoký stupeň polymorfizmu HLA systému zohledňují platné zásady pro označování HLA alel dané Světovou zdravotnickou organizací WHO (WHO nomenklatura). Princip je jednoduchý: Kaţdá alela je definována písemným označením

lokusu následovaným hvězdičkou (HLA-DRB1*), a poté kombinací 4 číslic (*0401), přičemţ první dvojčíslí určuje sérologickou specifitu dané alely, druhé pak označuje alelu na základě její aminokyselinové sekvence. Případné páté číslo charakterizuje tzv. "tichou" variantu alely, tzn. záměnu nukleotidů bez změny aminokyselinové sekvence.

Dědičnost HLA geny jsou děděny autozomálně kodominantně a vykazují mendelistický typ dědičnosti. Počet rekombinací v HLA systému je řídký, vyskytuje se přibliţně v 1 případů a častěji u ţen. Celá oblast od HLA-F aţ po HLA-DP se přenáší z rodičů na potomstvo jako haplotyp. V rámci rodiny se mohou vyskytnout teoreticky 4 různé kombinace rodičovských haplotypů, takţe sourozenci mohou být navzájem buď HLA identičtí, haploidentičtí (mají jeden haplotyp, v druhém se liší), anebo rozdílní. Rodiče jsou vůči svým dětem vţdy haploidentičtí [5]. Z genetického hlediska významný fenomén představuje existence vazebné nerovnováhy (linkage disequilibrium) v rámci HLA. Mnoho HLA genů se nalézá v tak těsné blízkosti, ţe se přenášejí z rodičů na potomky téměř vţdy společně. V důsledku této skutečnosti se v populaci vyskytují některé kombinace alel různých genů častěji, neţ by se očekávalo. Vazebná nerovnováha je významným faktorem v asociaci HLA antigenů s chorobami, protoţe mnohá onemocnění se v jejím důsledku váží s více antigeny.

Non-HLA geny Non-HLA geny jsou geny které se nepodílejí na základní funkci HLA systému. Z III třídy jsou to všechny, z II žádný a z I je to směs. Zjednodušeně můžeme říci, že geny které nejsou HLA jsou non-HLA. Tyto geny souvisejí též s funkcí imunitního systému, ne však vylučně s funkcí HLA.

lymfocyty bílá krvinka je leukocyt - typ bbílé krvinyky - T a B lymfocyty - specifická imunita - NK buňky nespecifická imunita - vznikají v z lymfatických kmenových buňek v kostní dřeni Aha takže lymfatické řečiště je více propustné proto to co nejde do cév jde sem pak se to odfiltruje a pak se to vrací do krevního řečiště.

KIR jsou na povrchu NK buňek a kde jsou teda NK buňky? NK je v podstatě lymfocyt a to je typ bílé krvinky. jo a nebudou teda spíš v lymfatické uzlině? leukocyty 1. granulocyty - neutrofilní, bazofilní a eozinogilní 2. agranulocyty - lymfocyty a monocyty

neutrofilní granulocyty jsou schopny vycestovat z kapilár do místa zánětu přeměněné monocyty přítomné v játrech v tělních dutinách (hrudní, břišní), ve slezině vy lymfatických uzlinách a kostní dřeni

KIR KIR jsou teda jak na HLA tak na non-HLA? Je to součástí genu - řadí se do přirozené (nespecifické) imunity narozdíl od B-buněk a T-buněk. - NK buňky představují 10-15% lymfocitů v periferní krvi - jsou to buňky

které reagují rychle a efektivně likvidují především nádorové buňky a buňky infokované virem

NK nemají antigenné specifické receptory, jak rozeznávají abnormální buňky? NK buňky identifikují molekuly vlastního MHC systému

jmenovitě HLA I. třídy, které jsou normálně exprimovány prakticky na všech buňkách v těle. Nádorové a některé virem napadené buňky potlačují expresi HLA I. třídy a tím se brání napadení cytotoxickými T lymfocyty (Restifo, 1993). Snížená exprese HLA I. třídy činí abnormální buňky citlivé k cytotoxicitě NK buněk (Karre, 1986). Molekuly HLA I. třídy rozpoznávají NK buňky pomocí pozitivních a negativních receptorů, které mohou inhibovat nebo naopak aktivovat NK buňky k „zabíjení“

Stručně lze shrnout, že NK buňky s potenciálem iniciovat cytotoxickou aloreakci používají KIR receptory jako inhibiční, směrem k „vlastním“, zdravým buňkám. Pokud však příslušný vlastní ligand HLA na cílové buňce chybí, pak dochází k iniciaci cytotoxické reakce. Proces interakce KIR/HLA a mechanismus regulace cytotoxicity NK buněk se jako

recptory imunoglobinové (protilátka - protein, který je schopen jako součást imunitního systému identifikovat a zneškodnit cizí objekty - bakterie a viry) v těle. Protilátky jsou nositeli humorální imunity. Jsou to krevní bílkoviny vznikající v mízní tkáni. povahy nacházejících se na povrchu Natural killers buněk a některých T-buněk (Variabilita v sekvenci).

KIR3D - prej tři skupiny ale to fakt divně popsaný (českej článek) něco s imunoglobulinovými doménami KIR2D

funkce KIR -

these genes are endcoded on chromosome 19. NK zabíjejí na základně interakce mezi KIR receptorem a HLA molekulou na povrchu buněk. Mohou mít různé podoby.

HLA i KIR jsou na různých chromozomech proto se segregují nezávisle a HLA schodni darci maji obvykle různé složení KIR genů

Struktura nukleových kyselin

jen skopírované z Nukleové kyseliny (polynukleotidy) jsou tvořeny dlouhými řetězci (mono)nukleotidů, vzájemně spojených fosfodiesterovými vazbami. Řadíme je k tzv.heteropolymérům, neboť jsou sestaveny z různých typů základních jednotek. Tato skutečnost je podstatná pro uchovávání a předávání informace, což je základní funkce nukleových kyselin v organismu. Homopolyméry (např. glykogen) obsahují pouze jeden typ monoméru (v našem případě glukózu), a tak nemohou plnit informační funkci.

2.9 Sekvence DNA

Je posloupnost písmen představující přimární strukturu reálné nebo hypotetické molekuly čí vlákna DNA, které má kapacitu nést informaci.

Používaná písmena A, C, G a T reprezentují čtyři nukleotidy ve vláknu DNA – adenin, cytosin, guanin a thymin, lišící se typem báze kovalentně vázané k fosfátové páteři. Posloupnost libovolného množství nukleotidů většího než čtyři lze nazývat sekvencí. Obvykle se sekvence vypisuje bez mezer, např. AAAGTCTGAC, ve směru 5 -> 3. Vzhledem k biologickým funkcím, které mohou záviset na kontextu, sekvence buďto mají anebo nemají smysl a jsou tedy kódující nebo nekódující DNA. Typem nekódující sekvence DNA je také tzv. „junk DNA“.

TO je z wiki bacha na to.

3 Sekvenační metody získávání DNA dat

Někdy se sekvenují pouze jisté části genomu které mají pro výzkumníka v daném okamžiku význam.

Sekvenování DNA je souhrný termín pro biochemické metody, jímiž se zjišťuje pořadí nukleových bází (A, C, G, T) v sekvencí DNA. Tyto sekvence jsou součástí dědičné informace v jádru. Adenin s thyminem a cytosins s guaninem.

zjišťvání přímární struktury nukleových kyselin (sekvencování)

Užitečné nejen ve výzkumu ale i v diagnostice nemocí či forenzní medicíně.

3.1 Porovnání vhodného dárce

V případě nepříbuzenských transplantací se vybírají potenciální dárci, kteří nemají s daným pacientem žádný děděný haplotyp. Snahou je najít takového dárce, který má shodné, přestože děděné od jiných rodičů, HLA antigeny. Informace o tom, jak jsou alely haplotypicky uspořádány obvykle chybí, proto je vždy nutná typizace maximálním rozlišením ve více HLA lokusech. Zjišťovaný minimální rozsah HLA shody se v jednotlivých transplantačních centrech liší. V současné době je u nepříbuzného páru požadována typizace vysokým rozlišením v lokusech HLA – A, B, C, DR a DQ (http://www.efiweb.eu/efi-committees/standards-committee.html). Pokud pacient a dárce mají stejné alely na všech těchto lokusech, hovoříme o shodě 10/10. Při jedné neshodě se jedná o shodu 9/10, při dvou o shodu 8/10. K typizaci se nejčastěji používá PCR – SSP (PCR se sekvenačně specifickými primery) či SBT (sequence based typing) technika, v posledních letech se stává zlatým standardem přímá sekvenace (SBT) HLA genu. fričová U HLA - A, B, C, DR a DQ požaduje se typizace v těchto lokusech. Pokud má dárce shodu ve všechn lokusech hovoříme o shodě 10/10 při jedné neshodě je to 9/10.
U Kir jsou 2 hlavní typy haplotyp A a B, které jsou definovány typem a počtem specifických KIR genů. Neexistuje žádné jednoduché univerzální kriterium definující a odlišující tyto haplotypy. Sekvenančí metody s elíší především rychlostí a cenou.

3.2 Sanger sequencing

K sekvenaci se použtívá gelová elektrofézy použitelná k sekvenování krátké sekvence jednovláknové DNA. využívá biologického procesu replikace DNA Vybraná sekvence se vloží do reakční směsi s radioaktivně označným primer

3.3 NGS next-generation sequencing

Je rychlé a relativné nenáročné zprácování jednotlivých vzorků. Tisíce až miliony sekvencí mohou být produkovány během jednoho sekvenčního procesu. K popularitě této metody nepomohla i komerciaze cenově dostupních stolních sekvenátorů.

3.4 Read

In DNA sequencing, a read is an inferred sequence of base pairs (or base pair probabilities) corresponding to all or part of a single DNA fragment. A typical sequencing experiment involves fragmentation of the genome into millions of molecules, which are size-selected and ligated to adapters. The set of fragments is referred to as a sequencing library, which is sequenced to produce a set of reads. Je to z wiki zase

V DNA sekvenování, read je odvozená sekvece párů bází odpovídající celému fragmentu DNA nebo jeho části To znamená že read je kus DNA který by mohl odpovídat nějakého konkrétnímu genu?

Pak tam ještě bylo psaný něco o read lenght Sekvenační technologie se liší? v délce vyrobenybch readů. Ready díkjy 20-40 párů bází (bp) jsou ultrakrátké Typická sekvenační metoda vytváří ready délky 100 až 500 bp

Sekvenančí platforma (Illumina) - podle toho se pak připravuje ta sekvenační knihovna

DNA knihovny - podle wikiskripta

DNA knihovny jsou kolekce klonovaných DNA fragmentů genomu určitého organismu (cDNA), které jsou skladovány uvnitř hostitelských organismů (zejména bakterií). cDNA (copy DNA, complementary DNA) je získávána přepisem z mRNA pomocí enzymu reverzní transkriptázy.

Kvalita knihovny Při přípravě sekvenční knihovny je důležité získat co nejvyšší úroveň složitosti. Jinými slovy, je důležité, aby konečná knihovna co nejvíce odrážela jedinečnost výchozího materiálu. Tento výsledek lze získat

především omezením počtu segmentových duplikací. Čím kratší jsou fragmenty, tím vyšší je pravděpodobnost, že jsou fragmenty méně specifické a mohou se zarovnat na více než jednom lokusu referenční sekvence. Složitost knihovny lze tedy v podstatě měřit procentem duplicitních čtení, které jsou přítomny v sekvenčních datech

READY - zase wikipedie In DNA sequencing, a read is an inferred sequence of base pairs (or base pair probabilities) corresponding to all or part of a single DNA fragment. A typical sequencing experiment involves fragmentation of the genome into millions of molecules, which are size-selected and ligated to adapters. The set of fragments is referred to as a sequencing library, which is sequenced to produce a set of reads

4 Analyza dostupných bioinformatických nástrojů pro zpracování NGS dat

4.1 ART

ART (next-generation sequencing read simulator) je sada simulačních nástrojů, které generují syntetické ready, jako kdyby byli získány sekvenováním pomocí NGS. Nástroj ART dokáže simulovat ready ze sekvenátorů Illuminas, 454 společnosti Roch a SOLid od společnosti Applied biosystém. Ready, vytvořené nástrojem ART jsou používány pro testování a analýzů nástrojů zpracovávající právě NGS sekvence jako například zarovávání (nástroj Bowtie).

ART je implementován v jazyce C++ a je dostupný s licencí GPL verze 3. Je dostupný pro operační systémy Linux, MacOs a Windows. Je možné ho používat i jako C++ package.

Data získána z FN Plzeň byla sekvenována nástrojem Illuminas proto i syntetické ready budou simulovat tento sekvenátor. Výstupy se čtou ve formátu FASQ a zarovnání ve formátu ALN. může generovat zarovánávání také ve formátu SAM nebo UCS BED. [2]

4.1.1 bordel

ART is freely available to public. The binary packages of ART are available for three major operating systems: Linux, Macintosh, and Windows. ART is also available as Platform-independent C++ source packages. Each package includes programs, documents and usage examples.

ART simuluje ready napodobobáním skutečných procesů sekvenování s empirickým chybovým modelem nebo quality profiles summarized from large recalibrated sequencing data ART může také simlovat čtené pomocí uživatelského vlastního read arror modelu nebo quality profiles

TODO - tohle úplně nechápu ART podporuje simulaci jedno párových, dvou párovcýh tří hlavních komerčních sekvenčních platfoem Výstupy se čtou ve formátu FASQ a zarování ve formátu ALN. ART může také genero-

vat zarovnávání ve formátu SAM nebo UCSC BED ART lze použít společně se simulátory varient genomů VarSim
to je odtud 454 sekvenování je pyrosekvenování, které cycklicky testuje přítomnost každého ze čtyř nukleotidů DNA (T, A, C, G)

SOLid ke kódování 16 různých dinukleotidů používá čtyči fluoresenční barevná barviva, každé barvivo kóduje čtyři dinukleotidy

tak jsem stáhla normálně nejnovější verzi z niehs.nih.gov a podle instrukcí co byli v souboru INSTAL dala

musí se brát v potaz že z toho generátoru nikdy nebudou data taková jako reálná.. realná budou horší

4.2 Bowtie

Bowtie je rychlý a paměťové efektivní nástroj pro zarovnávání krátkých sekvencí DNA na velké genomy. Indexace pomocí Burrows-Wheelere transformace dovoluje zarovnávání více než 25 milionů readů za CPU hodinu pro lidský genom s pamětí příblížně 1.3 gigabajtů. Bowtie přidáváa k Burrows-Wheeler technice backtracking algoritmus pro sledování nekonzistence. ??

4.2.1 Bordel

Bowtie je napsanej v c++ a používá knihovnu seqAn

Na lidském genomu je nástoj Bowtie v porovnání s nástoji Maq a SOAP rychlejší. Citlovost má bowtie srovnatelné s nástojem SOAP a o něco menší než Maq. Ale je možnost pomocí příkazové řádky zvýšit citlivost na úkor rychlosti běhu progamu. Oproti SOAP bowtie potřebuje méně paměti 1.3 GB RAM. Bowtie zarovnává 25 milionů readů za hodinu. může běžet paralelně.

indexi vytváří permanentní a lze je použít napříč běhy pro lidský genom je to 2.2 GB takže ho lze distribuovat přes internet rychlost a malá paměť způsobuje především Burrows wheeler v kombinaci s backtrackingem.

Podrporuje standardní vstupní formáty FASQ a FASTA.

Bowtie je open source.

na stránkách elixir-europe což je oragnizace co má dávat dohromady všechny vědecký veci a bla bla.

Tak tam je přímo Bowtie [5]

4.2.2 Bowtie 2

Note that SOAP2 and Bowtie do not permit gapped alignment of unpaired reads. memory footprint of Bowtie 2 (3.24 gigabytes) Bowtie 2 by mělo být

vhodnější pro delší ready než Bowtie1. We extracted a random subset of 1 million reads from each and aligned them with BWA-SW and Bowtie 2. We did not align with Bowtie, BWA or SOAP2 because those tools are designed for shorter reads. Bowtie už je překonanej nejenom Bowtie2 ale i BWA. Bowtie2 je podle studie znatelně lepší než Bowtie, SOAP2. tyhle výsledky jsou na syntetických readech

vypadá to že bowtie 2 už nepoužívá tamten index ale používá nějaký Full-text minute index–assisted search což vypadá že je kombinace burrows wheelera a ještě něčeho. We found that Bowtie 2, a method that combines the advantages of the full-text minute index and SIMD dynamic programming, achieved very fast and memory-efficient gapped alignment of sequencing reads

je zase open source [4]

šla jsem přes docker docker image ls - zobrazi vsechny image pak docker run a ID image sudo docker run -i -t 3c2b9a287f82 /bin/bash sudo docker ps -a

Tak jsem nakonec žádnej docker nepotřebovala a stáhla jsem to tady po kliknuti na bowtie binary release.

na strance 25.4 je řečeno o hledání tch nejlepších zarovnání a je tam možnost –best ale že je dvakrát nebo třikrát pomalejší než normální mod.. a jde o to že najde první přijatelný a to označní kdežto při tom best prohledá co nejvíc a hledá to nejlepší i mezi těma přijatelnýma a to je pomalý.

takže zarovnání by mohlo být teoreticky namapování na referenční gen???

Literatura

[1] Fryčová, M. Lze u pacientů s AML indikovaných k nepříbuzenské transplantaci provádět v klinické praxi výběr nepříbuzných dárců na základě KIR genotypů, 2016.

[2] Huang, W. et al. ART: a next-generation sequencing read simulator. 2012. Dostupné z: https://academic.oup.com/bioinformatics/article/28/4/593/213322.

[3] J, R. et al. *Nomenclature* [online]. Nucleic Acids Research, 2015. [cit. 2019/10/1]. 43:D423-431. Dostupné z: http://hla.alleles.org/misc/citing.html.

[4] Langmead, B. – Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. 2012. Dostupné z: https://www.nature.com/articles/nmeth.1923.

[5] Langmead, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. 2009. Dostupné z: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r25.

[6] MUDr. Pavel Jindra, P. D. *Imunopatologické a imunogenetické aspekty transplantací krvetvorných buněk a solidních orgánů.* PhD thesis, Universita Karlova v Praze, 2011.

[7] Robinson, J. et al. The IMGT/HLA Database. 2013. Dostupné z: https://www.ebi.ac.uk/ipd/index.html.

[8] Robinson, J. et al. IPD—the Immuno Polymorphism Database. 2013. Dostupné z: https://www.ebi.ac.uk/ipd/index.html.

[9] Smith, D. T. *Encyklopedie lidského těla.* 2005. ISBN 80-7321-156-4.